

## Logiciels anti-plagiat.



**Majid DACI**

### **0 - Position du problème :**

Avec l'avènement d'Internet et du tout numérique il est devenu banal de disposer facilement d'énormes quantités de texte concernant les domaines les plus divers de l'activité humaine.

Les fonctions copier et coller font partie de presque tout logiciel s'exécutent sur ordinateur.

Le copier coller est donc devenu tout naturellement l'un des premiers gestes de toute personne utilisant un ordinateur.

Une étude récente (sep. 2007) de l'Université de Lyon portant sur la problématique du plagiat a recueilli environ 1200 réponses dont 120 enseignants et plus de mille étudiants.

Quelques informations saillantes en ressortent :

#### **A - Le plagiat sur Internet**

- **4 étudiants sur 5** (79,7 %) déclarent avoir recours au "copier coller"
- **9 enseignants sur 10** ont déjà été confrontés au problème du "copier coller".

#### **B - Une confusion entre « citer » et « plagier » dans l'esprit des étudiants :**

- **3 professeurs sur 5** estiment que les citations sont mal identifiées dans les travaux des étudiants.
- Plus de 4 élèves sur 5 ont été formés à la « recherche documentaire ».

#### **C - Le passage au numérique**

- **4 travaux rendus par les élèves sur 5** existent sous forme numérique, mais 3 travaux sur 5 sont rendus sur papier ; d'où une difficulté de contrôle automatique par l'enseignant.

Par ailleurs « un commerce » de documents scolaires et universitaires a vu le jour en Amérique du Nord depuis de nombreuses années.

Ce commerce s'installe en France où des étudiants peuvent maintenant vendre et acheter des devoirs et rapports tout rédigés. (Cf [Le site academon](#) et [autres sites](#) ).

Il devient donc indispensable de disposer d'outils de détection du plagiat afin de faire face à cette situation.

L'objet de la présente étude est de recenser les principaux outils anti-plagiat et, dans la mesure du possible, donner leurs principales caractéristiques et ceci en effectuant une recherche sur Internet.

Il est à remarquer que ces outils servent aussi pour les webmestres des sites Internet pour s'assurer que leurs sites ne sont pas plagiés (pompés).

## I - Logiciels trouvés et catégorisation.

La recherche a permis de trouver sept outils **gratuits** et huit **payants**.

Ils se classent en quatre catégories selon la manière dont ils effectuent l'analyse de l'information à examiner :

- a) Les plus simples, ceux qui fonctionnent en faisant appel à des moteurs de recherche (Altavista, Google, MicroSoft Live, Yahoo,...), **sans installation de logiciel sur le poste** de l'utilisateur et sans se connecter au site de « l'éditeur ». C'est grâce à un Java script s'exécutant sur la page web que la recherche est effectuée.  
Ils acceptent trois sortes d'entrée possibles :
  - a1) recherche selon une liste de mots clef
  - a2) et/ou selon un document fourni par l'utilisateur (dans l'un ou l'autre des formats : Html, Word, texte)
  - a3) et/ou en recevant l'url d'une page web dont on souhaite examiner si elle est plagiée.Ils fournissent en retour une liste de pages contenant les mots clef ou présentant une similitude avec le document fourni ou la page Web.  
Ils peuvent aussi fournir un rapport de plagiat.
- b) Ceux qui, toujours **sans installation de logiciel sur le poste de l'utilisateur**, envoient le(s) document(s) à analyser vers le site de l'éditeur. L'examen se fait dans le site de l'éditeur vis à vis d'une base de documents propres à l'éditeur et/ou vis-à-vis d'Internet. En retour on reçoit ou bien une liste de documents présentant des similitudes et/ou une liste de sites plagiés/plagiants ainsi qu'un rapport de plagiat.
- c) ceux nécessitant **un logiciel à installer sur l'ordinateur local de l'utilisateur**. Ce logiciel effectuant la comparaison de deux ou plusieurs fichiers locaux entre eux et/ou utilisant un moteur de recherche Internet pour trouver des sites plagiés/plagiant; et ceci sans faire intervenir le site distant de l'éditeur lors de ces analyses. Pareil qu'en b), le retour est constitué d'une liste de sites plagiés

ainsi qu'un rapport de plagiat.

- d) ceux combinant les deux approches b) et c), c-à-d : **un logiciel à installer sur le poste local** qui envoie le document vers le site de l'éditeur. Le retour est identique à b) et c)

### **Outils gratuits ou ayant une version de base gratuite:**

- [CopyTracker](#) : Développé par des étudiants de l'**Ecole centrale de Lille**. Il gère les fichiers de type Pdf, html et Word en les comparant à un ensemble de fichiers ou une base de donnée locale ainsi qu'à des documents provenant d'une recherche Internet. Il fournit « un pourcentage de plagiat ». Nécessite Apache, Mysql et Filezilla.
- [WCopyfind](#) : En anglais. Il gère les fichiers de type html et Word. Il compare des fichiers locaux (pas d'Internet). Il indique les mots communs, le nombre de leurs occurrences et fournit un « pourcentage de plagiat ».
- [PlagiarismChecker](#) : Site de recherche de plagiat sur Internet, 32 mots clef max, moteurs : Google ou Yahoo, intégration possible avec Google-Alert, c-à-d la possibilité de recevoir automatiquement des emails d'alerte quant des nouvelles pages similaires sont détectées par Google dans les 10 ou 20 premiers résultats, [pour en savoir plus](#)
- [Copyscape](#) : la version gratuite détecte les pages plagiant une page web originale donnée avec ou sans citation de la page originale. TRES BON
- [TextGuard](#) : D'origine allemande, sans logiciel à installer, permet de rechercher les pages web plagiant (ou plagiées par) un texte fournit..
- [Plagiarismdetect](#) : Ce logiciel est en cours de développement par la Bauhaus-Universität Weimar. Il fonctionne via Internet et accepte les format texte et Word.
- [Articlechecker](#) : Fonctionne via Internet, accepte les formats texte et HTML (l'url d'une page web), peut comparer jusqu'à cinq document entre eux, utilise les moteur Google et Yahoo mais il sert de vitrine pour vendre d'autres logiciels

Tout le monde s'accorde à dire que « Les meilleurs outils gratuits de détection de plagiat sont les moteurs de recherche sur Internet »

### **Outils payants :**

- [Noplaiat.org](#) : Outil de détection de plagiat en ligne développé par des étudiants de la Faculté des Sciences de Luminy Marseille. Il gère les fichiers de type Pdf, html et texte brut. Actuellement en développement afin d'inclure de nouvelles fonctionnalités. L'outil était encore gratuit fin février ; il est devenu payant début mars 2008 sous le nom de [www.noplaga.com](#) .
- [Compilatio](#) ou [Pompotron](#) : Français, développé par la société de service [Six Degrés](#) Outil de détection de plagiat en ligne. Le document à examiner est envoyé sur le site de Compilatio qui le compare à sa base propre ainsi qu'à Internet. Le site garde une copie du document pour enrichir sa base
- [Euphorus](#) : Origine Pays –Bas Il offre quatre façon de travailler pour détecter le plagiat :

\* Ephorus Internet compare les travaux d'élèves avec des documents publiés sur

## Internet

- \* Ephorus Group exécute des comparaisons croisées entre les documents de groupes parallèles d'étudiants;
- \* Ephorus Database archive les travaux des élèves, permettant des comparaisons avec des travaux anciens ou d'autres facultés ;
- \* Ephorus Integrated est une solution complète de lutte contre le plagiat en intégrant Ephorus dans l'environnement numérique de travail de l'institution cliente. Cet outil est leader France.

- [EVE2](#): En anglais, il s'installe sur le PC et accepte les fichiers Word ou texte en entrée et les compare grâce à une recherche sur Internet.  
Peu cher (30 \$ la licence sans limitation de durée ni de nombre de documents traités), cet outil est très utilisé aux USA. Version d'évaluation 15 jours.
- [Turnitin](#) ou [iThenticate](#) : Le site se proclame comme étant la norme pour la prévention du plagiat en ligne. Le document à examiner est envoyé au site qui le compare à sa base locale contenant plus de 22 millions de textes, rapports, articles et comptes-rendus de toute provenance, avec lesquels il établit des comparaisons pour détecter les plagiat. Cette base s'enrichit par ailleurs au quotidien des 60.000 devoirs qui lui sont soumis pour vérification par les 6.000 institutions académiques de plus de 90 pays qui l'utilisent. Ce logiciel a été intégré aux plateformes : Blackboard et Moodle.  
Le service est facturé 0,80\$ par étudiant et par an.  
Les étudiants peuvent soumettre eux-mêmes leur travail au système Turnitin qui leur délivre un rapport ; le travail est rendu à l'enseignant accompagné de ce rapport.  
Cet éditeur a opère le site [www.plagiarism.org](http://www.plagiarism.org) qui voudrait laisser croire qu'il est indépendant de tout éditeur.  
Vu le nombre considérable de documents dans la base, les étudiants éprouvent des réelles difficultés pour produire des textes qui ne soient pas « taggés » plagiat, d'où la nécessité de re soumettre.
- [Scriptum](#) ou [CatchItFirst](#) : Anglais canadien. Cet outil en ligne est offre la fonctionnalité d'espace d'échange de travaux entre étudiants et enseignants ;et on a l'impression qu'accessoirement le travail de l'étudiant est examiné contre le plagiat vis-à-vis d'Internet. Les formats acceptés ne sont pas spécifiés ; mais les documents déposés sont systématiquement convertis en Pdf. L'enseignant peut directement annoter et **noter** le travail de l'étudiant qui peut re-soumettre son travail
- [Doccop](#) : Outil en ligne en anglais ; il gère des documents Word et Pdf en les comparant à Internet; il ne garde pas de copie de l'original.
- [PlagiarismFinder](#) : Origine allemande, il s'installe sur le PC et accepte les formats Html, Word, Pdf, texte et, Rtf. Coût d'une licence 125\$ sans limitation de durée ni de nombre de documents traités. Version d'évaluation bridée 30 jours.

## II – Critères de comparaison :

Devant la liste d'outils ci-dessus, on se pose la question de trouver des critères « les plus pertinents possibles » permettant de les comparer dans l'objectif de les utiliser dans le contexte de l'école.

L'essentiel des critères résulte des considérations suivantes :

- 1) La disponibilité du logiciel (ou de l'interface) en français constitue a priori un critère de bonne adéquation pour l'école.
- 2) Les formats supportés (Word, Html, Pdf, texte, Rtf...)
- 3) Le coût – quand il est précisé par l'éditeur - est un critère important,
- 4) La production de rapport avec un pourcentage quantifiant le plagiat (%) et possibilité de mettre en face à face (fàf) les documents plagiant et plagié.
- 5) La présence de la fonctionnalité de comparaison de deux ou plusieurs fichiers entre eux et dans des formats différents (par ex l'un Word, l'autre Pdf,...) et sans recherche sur le net,
- 6) Les outils ne gardant pas une copie du document à analyser respectent théoriquement mieux les droits d'auteur que ceux qui systématiquement s'autorisent à en garder une.
- 7) La facilité d'installation, d'utilisation et ergonomie,
- 8) Disponibilité d'une version d'évaluation permettant de tester le produit ; les autres ne peuvent être évalués qu'au travers d'avis glanés sur le net.
- 9) La base de documents propres située sur le site de l'éditeur est formée essentiellement de documents en anglais (EN) ou aussi en français (FR)
- 10) Références en France et/ou dans les pays francophones,
- 11) L'appréciation de l'efficacité du moteur de recherche propre utilisé par l'outil quand cela est possible,
- 12) La possibilité donnée par certains outils de permettre à l'élève de se corriger et resoumettre son document,
- 13) Quand l'analyse s'effectue à distance, la possibilité d'utiliser le site de l'éditeur comme espace d'échange entre élèves et enseignants,
- 14) Disponibilité d'autres fonctionnalités, à préciser.

Cependant, deux critères importants ne figurent pas dans cette liste : **la robustesse du logiciel** ainsi que **la pérennité** du fournisseur.

© Original Artist  
Reproduction rights obtainable from  
www.CartoonStock.com



'YOU'VE COPIED ALL THIS OFF  
THE INTERNET...'

### III - Indications sur Le « matériel » utilisé pour examiner les logiciels antiplagiat :

- Quand il s'est agit de trouver les pages sur Internet plagiant un texte donné, j'ai utilisé les textes suivantes :

1 - « L'École des ponts est une grande école française qui offre la totalité des formations composant le référentiel international actuel des diplômes, caractéristique du cadre de haut niveau à profil scientifique et technique. Cette offre est harmonisée au système LMD (licence - master - doctorat) adopté par 40 pays dans le cadre de l'espace européen de l'enseignement supérieur. » Extrait du [site de l'école](#)

2 – « The internationally recognized opera singer served as a United Nations Messenger of Peace for nearly a decade before his death at age 71 in Modena, Italy on 6 September 2007. Mr. Pavarotti was strongly committed to alleviating the suffering of children in war-torn countries and generated millions of dollars in humanitarian aid. » Extrait du site de [l'ONU sur Pavarotti Messenger de la paix](#)

3 – « Tout logiciel disponible pour téléchargement à partir de ce serveur (le "Logiciel") est protégé par les droits d'auteur de Microsoft et/ou de ses fournisseurs. L'utilisation du Logiciel est régie par les termes du contrat de licence utilisateur final, s'il existe, qui est inclus dans le Logiciel ou qui l'accompagne (le « Contrat de licence »). L'utilisateur final doit accepter les termes du Contrat de licence pour pouvoir installer le Logiciel. » Extrait des clauses de [Copyright du site Microsoft France](#)

4 – J'ai aussi soumis plusieurs documents Word et PDF de ma propre rédaction.

- La majeure partie des outils offre des versions d'essai ou propose un nombre d'essais d'analyses gratuites (entre 3 et 10 environ) permettant ainsi de se faire une idée du fonctionnement de l'outil.

Les 2 tableaux suivants présentent les outils, gratuits et commerciaux, selon ces critères :

## VI – Analyse des outils gratuits:

### Analyse des outils gratuits

Outils	Versi on FR	Formats supportés	Logiciel / Internet	Rapport de plagiat	Compar ais, de fichiers	Appro priat, doc,	Ergono mie a-b- c-d-	Format du résultat	Base en EN/FR	Réfs en Franc e	Moteur propre	Resou mettre	Elèves + enseigs	Autres fonctionn,	Remarques
<a href="#">CopyTracker</a>	oui	Word, HTML, PDF, Texte	logiciels + Internet	oui % + fàf	oui	non	b	Html, PDF PPT Doc	sans		non	oui	non	* Gestion fichiers locaux * impression Rapport PDF	
<a href="#">WCopyfind</a>	non	Word, HTML, texte	logiciel	oui % en html	oui	non	b	Gratuit	sans		non	oui	non		Pas de recherche Internet
<a href="#">Copyscape</a>	non	HTML	Internet	Oui fàf + sites	non	non	b	10/mois	sans		oui	non	non	Fournit les 10 pages plagiantes	Format des plagiantes : PDF/HTML
<a href="#">TextGuard</a>	non	Texte, HTML	Internet	sites	non	non	c	Html, PDF	sans		oui	oui	non		Interface en allemand
<a href="#">Plagiarismdetect</a>	non	Texte, Doc	Internet	% + sites	non	non	a	Html	sans		non	oui	non	2 niveaux : analyse normale ou poussée	
<a href="#">Plagiarismchecker</a>	non	Texte, page web	Via Internet	Liste de sites	non	non	b		non	?	non	oui	non		Un site simplifiant Google
<a href="#">Articlechecker</a>	non	Teste, HTML	Internet	oui	oui	?	c	Html	?	?	non	oui	non	Semble servir de vitrine pour vendre autres produits	A éviter

## V – Analyse des outils payants :

## Analyse des outils payants

Outils	Versio n FR	Formats supportés	Logiciel / Internet	Rapport de plagiat  Oui, mais2 %# pour un même documt.	Compara, de fichiers	Stock/Ap prop, doc,	Ergon omie	Coût (par analyse)	Base en EN/FR	Réfs en France	Moteur propre	Resou mettre	Elèves + enseigns	Autres fonctio,	Remarques
<a href="#">Noplagiat.com</a>	oui	Texte HTML PDF	Via Internet		?	?	?	(De 1,20 à 0,60€)	?	Existe	?	non	non		Payant depuis le 01/03/08 Semble s'accaparer le marché national Mutualisati on des bases des écoles
<a href="#">Compilatio</a>	oui	Texte HTML	Via Internet	oui	?	?	?	(1,30€ et Abont 150€/an)	?	oui	?	non	non	100Megas , Uniqmt en français L'école demeure propr. du doc	
<a href="#">Euphorus</a>	oui	Texte, Html, Word, PDF	Via Internet	Oui % fàf Oui en	oui	Oui/non	a	1,6€/élè/an pour 650 élèves	oui	oui	oui	oui	oui		
<a href="#">EVE2</a>	oui	Texte, Html, Word	logiciel	RTF	non	non	b	30\$	non	oui	non	non	non		
<a href="#">Turnitin</a>	oui	Texte, Html, Word, PDF	Via Internet	Oui % fàf en Html	oui	oui	b	(0,80\$/élève/an )	Oui EN	oui	oui	oui	oui	<a href="#">Version Moodle intégré</a>	Moodle hébergé 5\$ /élève/an
<a href="#">Scriptum</a>	non	Texte, Html, Word, PDF	Via Internet	Oui % et ++ en PDF	oui	non	a	<a href="#">120\$/an pour 1- 50 élèves</a>	non	?	oui	oui	oui	Mini plateforme Réglage	
<a href="#">PlagiarismFinder</a> :	non	Texte, Html, Word, PDF	Via Internet	Oui % fàf en Html	oui Nb. illimité de fichiers	?	d	125\$ : usage illimité Free up to 7 files of 250Kwords each	non	?	?	oui	non	de la finesse de l'examen 250Kmots = 125 pages = un livre	Examen en 4 étapes : fastidieux Peut comparer nb illimité de rapports
<a href="#">Doccop</a>	non	Texte, Word,PDF	Via Internet	Oui % fàf	entre eux	non	a		non	?	NA	oui			

% : pourcentage

# : différents

fàf : face à face



## IV – Conclusion :

Deux outils gratuits ([TextGuard](#) et [Articlechecker](#)) et un outil payant ([PlagiarismFinder](#)) ont présentés des disfonctionnements sérieux pendant les essais.

J'ai pu tester tous les autres.

Pour classer ces produits, la méthodologie aurait été d'attribuer un poids à chaque critère et de comparer les sommes des poids obtenues par les outils.

Mais pour quelques produits, je n'ai pas réussi à trouver sur Internet les réponses à quelques critères (là où il y a un ?).

Cependant, aussi bien parmi les gratuits que les payants, des produits sont sortis du lot, aussi bien lors des tests qui se sont bien déroulés que par des critères de bonne facture.

### **Pour les gratuits :**

- [CopyTracker](#) me semble arrivé en tête et ce malgré le fait qu'il nécessite l'installation locale de PHP, MySQL et Apache ; il est français et offre une grande richesse de fonctionnalités (formats supportés, rapport de plagiat..)
- Le second serait [plagiarismdetect](#) pour sa simplicité, sa rapidité et sa souplesse

### **Pour les payants :**

- [Euphorus](#) est un excellent produit ; tous ses critères sont bons et j'ai reçu une cotation pour l'école (1,6€/élève/an pour 650 élèves).
- Il est suivi de près par [Turnitin](#) et [Compilatio](#)  
Ces trois produits ont des versions françaises et des représentations en France. Par ailleurs ils satisfont parfaitement les deux importants critères de pérennité et de robustesse.